Protein Isoelectric Point as a Predictor for Increased Crystallization Screening Efficiency

Katherine A. Kantardjieff [1] and Bernhard Rupp[2,3]

[1] *Department of Chemistry and Biochemistry and W.M. Keck Foundation Center for Molecular Structure, California State University Fullerton, Fullerton, CA 92834-6866, USA*

[2] *Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843-2128, USA*

[3] *Biology and Biotechnology Research Program, L-448, Lawrence Livermore National Laboratory Livermore, CA 94551, USA*

Correspondence: Phone: (714) 278-3752, Fax: (734) 939-4225

E-mail: kkantardjieff@fullerton.edu, br@llnl.gov

Running title: pI as a crystallization predictor

Keywords: pI, isoelectric point, protein crystals, crystallization pH, high-throughput crystallography

Abstract

Motivation: Increased efficiency in initial crystallization screening reduces cost and material requirements in structural genomics. Because pH is one of the few consistently reported parameters in the Protein Data Bank (PDB), the isoelectric point, pI, of a protein has been explored as a useful indirect predictor for the optimal choice of range and distribution of the pH sampling in crystallization trials.

Results: We have analyzed 9596 unique protein crystal forms from the August 2003 Protein Data Bank and have found a significant relationship between the calculated pI of successfully crystallized proteins and the difference between pI and reported pH at which they were crystallized. These preferences provide strong prior information for the design of crystallization screening experiments with significantly increased efficiency and corresponding reduction in material requirements, leading to potential cost savings of millions of US$ for structural genomics projects involving high throughput crystallographic structure determination.

Availability: A prototype example of a screen design and efficiency estimator program, CrysPred, is available at http://www-structure.llnl.gov/cryspred/ .

Contact: br@llnl.gov

Introduction

One of the key components in for any high throughput X-ray crystallography (HTPX) project is an efficiently operating crystallization facility. In the absence of any predictive *ab initio* algorithms or rules for crystallization, an optimized crystallization screening protocol should maximize the probability of successes while minimizing the number of chemical components, general physical parameters, and method specific parameters to be sampled, leading to increased throughput at reduced cost [1].

Minimizing the amount of protein sample used for crystallization screening is a major goal in any HTPX effort. It can be achieved by miniaturization using various nanotechnologies [2-4] and/or by increasing the crystallization success rate. Various screening protocols and methods searching the protein crystallization space for successes are currently used. While there is complete agreement on the desirable results (leading towards diffracting crystals from which a structure may be obtained) the best way to achieve such results, however, has been hotly debated. Although a number of studies have attempted to provide improved crystallization strategies [5], much of the 'knowledge' disseminated about protein crystallization continues to be anecdotal, with little statistical evidence or control experiments to prove its general efficiency or usefulness. Considering the wide variety of physical, chemical and method related parameters, very few parameters are sampled (and reported) with sufficient overlap to allow their direct use as a predictive means for optimizing crystallization success [6]. One parameter that is however frequently reported, regardless of the crystallization strategy employed, is the pH of the crystallization cocktail. Although the pH is rarely measured or accurately determined in a crystallization experiment, its use as a predictor for crystallization success, either globally or in correlation with the minimum solubility of a given protein at its isoelectric point, pI, appears attractive. Unfortunately, no direct correlation between minimum solubility at the pI and the pH of crystallization has ever been established.

As the pH is one of the few consistently reported parameters in the Protein Data Bank (PDB), we have analyzed 9596 unique protein crystal forms from the August 2003 Protein Data Bank and have found a significant relationship (not a direct correlation) between the calculated isoelectric point, pI, of successfully crystallized proteins and the reported pH at which they were crystallized. Specifically, there is a clearly preferred range of crystallization pH for acidic and basic proteins, and these preferences provide strong prior information for the design of crystallization screening experiments of significantly increased efficiency. An overall efficiency increase of 30 to 50% compared to random pH screening in protein crystallization and corresponding reduction in material requirements could lead to cost savings of millions of US$ for structural genomics projects using high throughput crystallographic structure determination.

System and Methods

We have used the SEQRES records of 9596 PDB entries comprising a nonredundant protein data set [7] which contain the sequence of the entire expressed construct including any tags, fusions or linkers, to calculate the pI using the pK_a values of Bjellqvist et al. [8], and we have treated complexes of proteins and nucleic acids (469 entries) as a separate group. The frequency distribution for pI of proteins is biomodal (Figure 1A), with highest frequencies (modes) at approximately pH 5.7 and 9.0, similar to the pI distribution seen for proteins encoded by sequenced genomes. (See for example [9-11].) The frequency distribution for reported crystallization pH of proteins is unimodal, with mean = 6.7, median = 6.9 and mode = 7.5 (Figure 1B). For the complexes, we observe a similar bimodal distribution of pI, with modes at 6.1 and 9.5 (Figure 1C), and a unimodal distribution of crystallization pH, with mean = 6.6, median = 6.5, mode = 6.5 (Figure 1D). A similar distribution of crystallization pH has been observed from successful crystallizations of proteins resulting from unbiased random screening experiments in a structural genomics initiative [6].

We find that while there is no statistically significant direct correlation between the pI of crystallized protein and pH of crystallization, there is a good correlation ($R^2 = 0.62$) between the pI of crystallized protein and the *difference* between pH of crystallization and pI (Figure 2). The delta (pH – pI) histograms for acidic and basic proteins are shown in Figure 3. It is apparent that acidic proteins crystallize with highest likelihood ~0-2.5 pH units *above* their isoelectric point, whereas basic proteins preferably crystallize ~0.5-3 pH units *below* their isoelectric point. Extreme values of pH do not contribute significantly to successful crystallization for most proteins, except for those that have unusually high or low pI values. For nucleic acid-bound proteins (not shown), the correlation is also strong ($R^2 = 0.77$), with similar tendencies for optimal pH of crystallization, ~0-2 pH units *above* the pI for acidic proteins, ~2-4 pH units *below* the pI for basic proteins. We have not accounted for the pI of DNA (pH ~ 4), however, which generally lacks functional groups that change ionization state near physiological pH [12]. Although conditions for crystallizing DNA-protein complexes have been shown to be similar to protein-only crystallization conditions, we do not use this last correlation for predictive purposes, due to the above mentioned uncertainties, as well as the limited number of data points.

Implementation

To demonstrate the utility of our analysis, we have implemented a prototype pH range calculator, CrysPred (http://www-structure.llnl.gov/cryspred/). The purpose of this small server-based applet is to show how prior information can be used to optimize efficiency of initial crystallization screening in HTPX. Effective initial crystallization screening aims to identify with the highest *overall* efficiency (least material, supplies and resources, and thus cost) the proteins that are most likely to yield useful or suitable crystals and structures. The purpose of efficient initial screening is *not* to find conditions for each and every protein, but to focus resources (scale-up, Se-Met incorporation, etc) on those proteins which have the highest probability to yield structures with the least effort (a.k.a. 'the first cut', 'cherry picking', etc).

CrysPred accepts as input the amino acid sequence of the protein moiety to be crystallized, including the sequence of any tags, linkers or fusions, if present, and the number of crystallization experiments to be attempted. The program returns the calculated pI for the protein, as well as a histogram showing the "delta" bins (pH-pI) for successfully crystallized proteins with similar pI, grouped in clusters of two pH units. A table is provided also, showing the delta bin frequency expressed as a percentage of the pI cluster, the population of experiments (equal distribution) for a random screen, the recommended population of experiments based on the "delta" prior information, and a suggested range of pH for the specified experiments (Figure 4). Finally, CrysPred estimates the expected efficiency increase compared to pH screening with equally populated bins of each pH over the selected range. Depending on the shape of the corresponding frequency distribution and the extent of the pH sampling range, the total savings of material is predicted typically to be between 30-50 %.

The values from CrysPred can be easily imported into any customizable screen generator that allows to define the frequency of occurrence for selected pH ranges (for example, CrysTool [13, 14]). The pH frequency distribution data are available for download from the CrysPred site to allow a custom implementation if desired.

Discussion

Methods for choosing protein crystallization conditions have largely been empirical, based on knowledge of what has worked in the past [15]. More recently, random screening methods have been developed [13, 14], and it is anticipated that statistical analysis will provide predictive frameworks that increase the probability of producing high quality crystals. Because pH is one of the few consistently reported crystallization parameters in the Protein Data Bank (PDB), we have completed such a statistical analysis and implemented into a predictive framework called CrysPred the significant relationship between calculated isoelectric point, pI, of successfully crystallized proteins and the reported pH at which they were crystallized.

Crystallization is a special case of phase separation from a thermodynamically metastable solution under the control of kinetic parameters [6]. While control over kinetic parameters such as nucleation or growth rates is rather difficult to achieve, attractive interaction between molecules as a thermodynamically necessary – but not sufficient – condition for crystallization can be discussed on the basis of thermodynamic excess properties, in particular their manifestation in the second virial coefficient, B_{22}, as determined by static light scattering and osmotic pressure measurements.

More than fifty years ago, Zimm examined theoretically the osmotic second virial coefficient of proteins, B_{22} [16]. At the molecular level, B_{22} reflects the nature of protein-protein interactions, which involve van der Waals attractions, electrostatic repulsions, noncentrosymmetric dipole interactions, hydrophobic interactions, hydrogen bonding and ion bridge mechanisms. More negative values of $B_{22,}$ are indicative of more attractive interactions. Protein solubility is affected by solvent and additives, which alter protein size and surface characteristics [17]. Quantitative links between the second viral coefficient and solubility have suggested that large classes of globular proteins will exhibit similar solubility with the same normalized B_{22} [18-20]. A number of groups [17, 21-26] have shown that, for proteins under conditions where they were crystallized, the second virial coefficient is negative, falling in a narrow range termed the “crystallization slot” [21] , and it is well documented that protein crystallization occurs in or close to attractive regimes [26]. Tardieu et al. have recommended that to crystallize soluble proteins (starting from a monodisperse solution), one should start far from precipitation and gently adjust repulsive interactions towards more attractive ones [26]. However, although interactions tend to be attractive near the pI, in accord with the van der Waals potential, van der Waals forces are considerable only for small compact proteins [26].

A number of studies on protein solutions and crystals [27-30] have shown that protein-protein interactions can be described by a sum of surface contacts between proteins, but that the mutual arrangement of proteins requires some anisotropy [24, 27, 31] or complementarity (molecular recognition) [32]. Neal et al. have shown that as pH values approach pI, and charge and repulsive interactions are decreased, B_{22} becomes more negative at constant values of ionic strength [32]. The magnitude of repulsive interactions and appearance of attractive interactions depend on the spatial distribution of charges and not simply on the global net charge of the protein, although accounting for short-range effects due to hydrogen bonding and solvation is not straightforward. Whereas changing the pH to approach the isoelectric point reduces the overall protein charge and decreases longer range electrostatic repulsion, Debye-Hückel screening of repulsive charge interactions may be exploited by searching for crystals under conditions of pH away from the pI [33]. B_{22} (and the possibility to crystallize) is determined largely by relatively few attractive interactions, the molecular configurations of which are influenced by pH and ionic strength.

Thus, while buffering at a pH equal or very near to the pI value of a protein offers a reasonable probability of yielding crystals, this pH is not necessarily that value with the highest probability. The “knowledge” occasionally perpetuated at protein crystallization workshops and by unreviewed publications that a protein has the best chance of crystallizing at a pH very near its solubility minimum, pI, is not reflected statistically in the PDB data. We have found one commercially available crystallization screen that recommends empirically, "The high efficiency of this kit can be further improved by pre-determining the isoelectric point (pI) of the subject macromolecule, followed by screening within a range at or near that value (within 2-3 pH units of the pI)." [34] Our statistical analysis suggests optimal pH ranges for crystallization screening and, to improve efficiency of any crystallization screen, we recommend that the pI of the protein moiety to be crystallized be used to design an optimized pH distribution for incorporation into screening experiments.

Acknowledgements

LLNL is operated by University of California for the US DOE under contract W-7405-ENG-48. This work was funded by NIH P50 GM62410 (TB Structural Genomics) center grant and the Robert Welch Foundation. The W.M. Keck Foundation Center for Molecular Structure is a core facility of the California State University Program for Education and Research in Biotechnology.

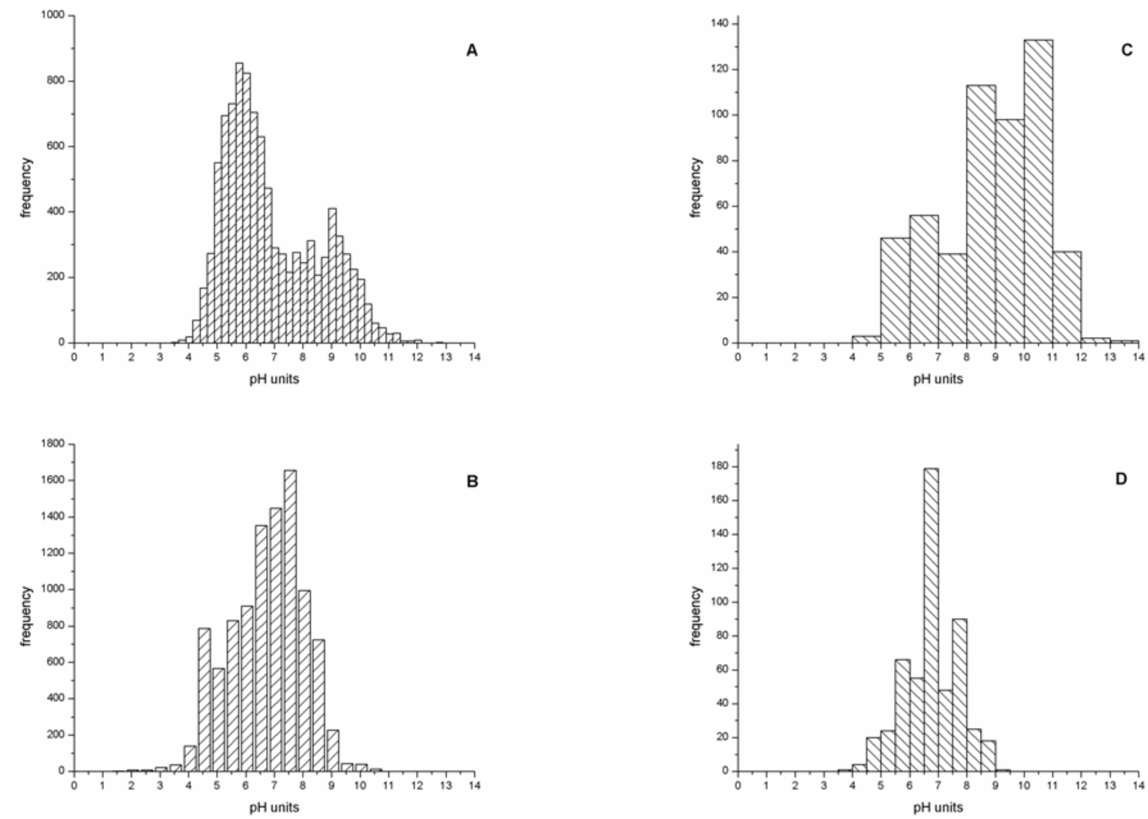

Figure 1 Frequency distributions. (A) pI of successfully crystallized proteins. (B) reported pH of crystallization for proteins. (C) pI of successfully crystallized protein-nucleic acid complexes. (D) reported pH of crystallization for protein-nucleic acid complexes.

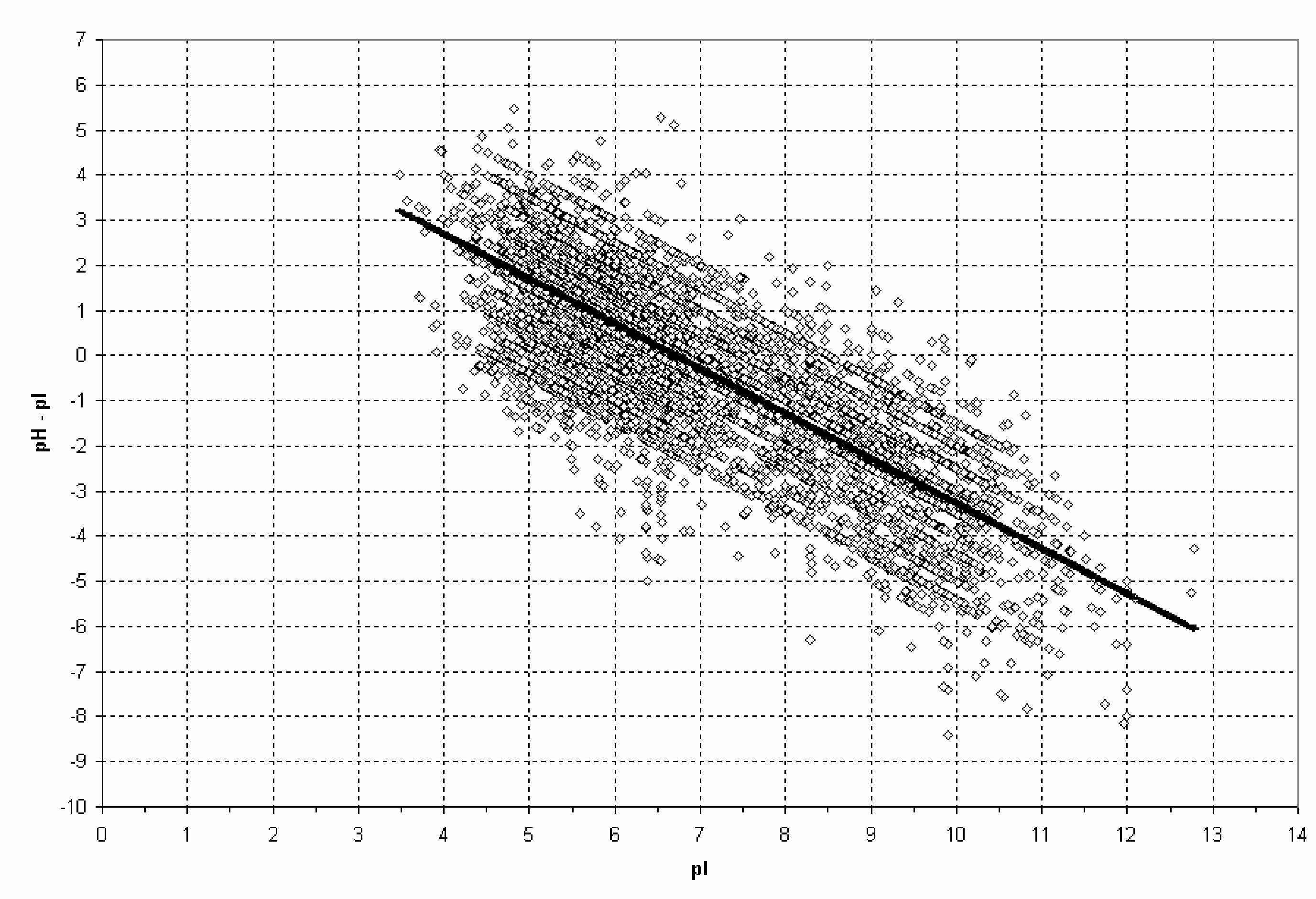

Figure 2 Correlation between pI and pH. Correlation between calculated pI of successfully crystallized protein and difference between reported crystallization pH and pI. $R^2 = 0.62$, P-value < 10^{-7}. Not shown, protein-nucleic acid complexes ($R^2 = 0.77$, P=value < 10^{-7})

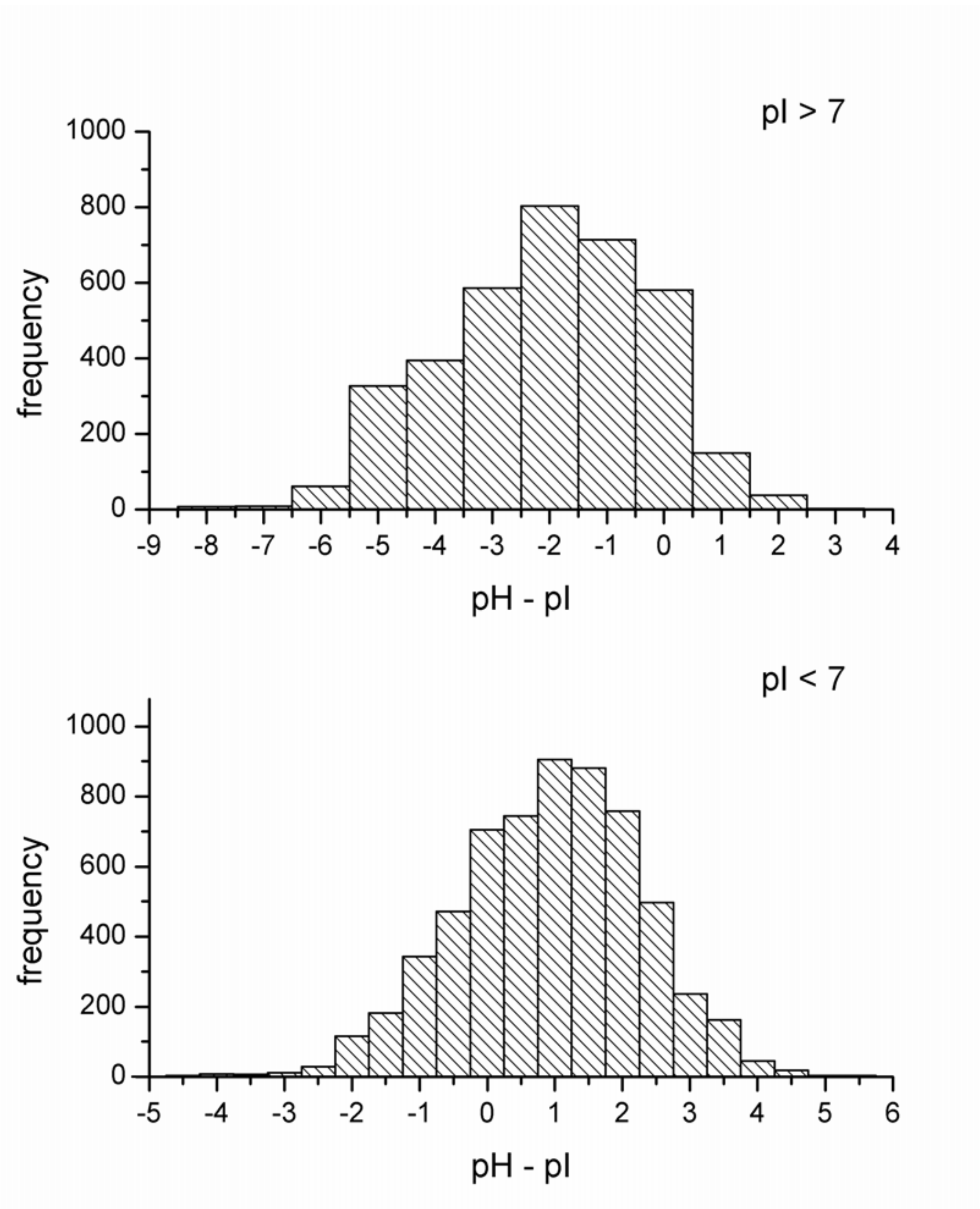

Figure 3 Delta histograms for successfully crystallized proteins. Top panel shows frequency distribution of the difference between crystallization pH and pI of successfully crystallized basic proteins. Bottom panel shows this frequency distribution for acidic proteins. It is clear that basic proteins have a tendency to crystallize 0.5-3 pH units *below* their pI, whereas acidic proteins prefer to crystallize 0-2.5 pH units *above* their pI. Similar tendencies are observed for protein-nucleic acid complexes, although this is shifted 2-4 units *below* the pI for basic proteins.

Estimated pI = 6.15

```
Table for cutoff excluding bins with expected success rates below 0.1%
pH-pI bin  : -8.0 -7.0 -6.0 -5.0 -4.0 -3.0 -2.0 -1.0  0.0  1.0  2.0  3.0  4.0  5.0  6.0
Expected % :  0.0  0.0  0.0  0.0  0.0  0.3  1.1  6.2 14.6 25.6 28.9 19.0  3.8  0.5  0.0
Population of 288 experiments in 9 bins :
equal pop. :    0    0    0    0    0   32   32   32   32   32   32   32   32   32    0  288
suggested  :    0    0    0    0    0    1    3   17   41   73   83   54   10    1    0  288
Expected relative hit rates
equal pop. :  0.0  0.0  0.0  0.0  0.0  0.1  0.4  2.0  4.7  8.2  9.3  6.1  1.2  0.2  0.0 32.0
suggested  :  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.1  6.1 18.9 24.1 10.4  0.4  0.0  0.0 61.1
pH            ---  ---  ---  ---  ---  3.1  4.1  5.1  6.1  7.1  8.1  9.1 10.1 11.1  ---
Experiments:  ---  ---  ---  ---  ---    1    3   17   41   73   83   54   10    1  ---
Expected efficiency increase compared to pH screening with equally populated bins: 91%


Table for cutoff excluding bins with expected success rates below 1.0%
pH-pI bin  : -8.0 -7.0 -6.0 -5.0 -4.0 -3.0 -2.0 -1.0  0.0  1.0  2.0  3.0  4.0  5.0  6.0
Expected % :  0.0  0.0  0.0  0.0  0.0  0.0  1.1  6.2 14.6 25.6 28.9 19.0  3.8  0.0  0.0
Population of 288 experiments in 7 bins :
equal pop. :    0    0    0    0    0    0   41   41   41   41   41   41   41    0    0  287
suggested  :    0    0    0    0    0    0    3   17   42   74   84   55   10    0    0  288
Expected relative hit rates
equal pop. :  0.0  0.0  0.0  0.0  0.0  0.0  0.5  2.5  6.0 10.5 11.9  7.8  1.5  0.0  0.0 40.6
suggested  :  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.1  6.2 19.0 24.3 10.5  0.4  0.0  0.0 61.6
pH            ---  ---  ---  ---  ---  ---  4.1  5.1  6.1  7.1  8.1  9.1 10.1  ---  ---
Experiments:  ---  ---  ---  ---  ---  ---    3   17   42   74   84   55   10  ---  ---
Expected efficiency increase compared to pH screening with equally populated bins: 51%


Table for cutoff excluding bins with expected success rates below 2.0%
pH-pI bin  : -8.0 -7.0 -6.0 -5.0 -4.0 -3.0 -2.0 -1.0  0.0  1.0  2.0  3.0  4.0  5.0  6.0
Expected % :  0.0  0.0  0.0  0.0  0.0  0.0  0.0  6.2 14.6 25.6 28.9 19.0  3.8  0.0  0.0
Population of 288 experiments in 6 bins :
equal pop. :    0    0    0    0    0    0    0   48   48   48   48   48   48    0    0  288
suggested  :    0    0    0    0    0    0    0   18   42   75   85   55   11    0    0  287
Expected relative hit rates
equal pop. :  0.0  0.0  0.0  0.0  0.0  0.0  0.0  3.0  7.0 12.3 13.9  9.1  1.8  0.0  0.0 47.0
suggested  :  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.1  6.2 19.2 24.6 10.6  0.4  0.0  0.0 62.2
pH            ---  ---  ---  ---  ---  ---  ---  5.1  6.1  7.1  8.1  9.1 10.1  ---  ---
Experiments:  ---  ---  ---  ---  ---  ---  ---   18   42   75   85   55   11  ---  ---
Expected efficiency increase compared to pH screening with equally populated bins: 32%


Table for cutoff excluding bins with expected success rates below 5.0%
pH-pI bin  : -8.0 -7.0 -6.0 -5.0 -4.0 -3.0 -2.0 -1.0  0.0  1.0  2.0  3.0  4.0  5.0  6.0
Expected % :  0.0  0.0  0.0  0.0  0.0  0.0  0.0  6.2 14.6 25.6 28.9 19.0  0.0  0.0  0.0
Population of 288 experiments in 5 bins :
equal pop. :    0    0    0    0    0    0    0   57   57   57   57   57    0    0    0  285
suggested  :    0    0    0    0    0    0    0   18   44   78   88   58    0    0    0  287
Expected relative hit rates
equal pop. :  0.0  0.0  0.0  0.0  0.0  0.0  0.0  3.5  8.3 14.6 16.5 10.8  0.0  0.0  0.0 53.7
suggested  :  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.2  6.5 20.0 25.6 11.0  0.0  0.0  0.0 64.3
pH            ---  ---  ---  ---  ---  ---  ---  5.1  6.1  7.1  8.1  9.1  ---  ---  ---
Experiments:  ---  ---  ---  ---  ---  ---  ---   18   44   78   88   58  ---  ---  ---
Expected efficiency increase compared to pH screening with equally populated bins: 19%
```

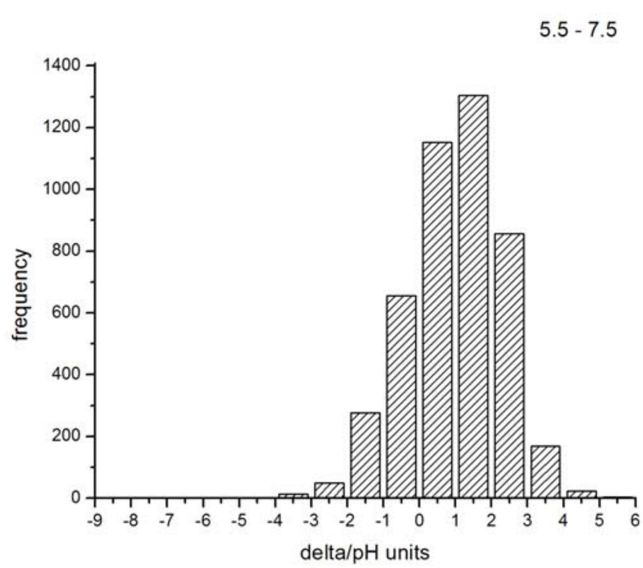

Figure 4 Sample CrysPred output. Shown are the calculated pI for the protein and a histogram of the "delta" bins (pH-pI) for successfully crystallized proteins with similar pI, grouped in clusters of two pH units. Table reports the delta bin frequency expressed as a percentage of the pI cluster, the equal population of experiments for a random screen, the recommended population of experiments based on the pH prior information, and a suggested range of pH for the specified experiments. Expected efficiency increase compared to pH screening with equally populated bins of each pH over the selected range is also predicted (in this example between 19 and 92%).

References

1. Rupp, B., *High throughput crystallography at an affordable cost: The TB Structural Genomics Consortium crystallization facility.* Acc Chem Res, 2003. **36**: p. 173-181.
2. Bodenstaff, E., F. Hoedemaker, E. Kuil, H. deVrind, and J. Abrahams, *The prospects of nanocrystallography.* Acta Crystallogr, 2002. **D59**: p. 1901-1906.
3. Hansen, C., E. Skordalakes, J. Berger, and S. Quake, *A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion.* Proc Natl Acad Sci USA, 2002. **99**: p. 16531-16536.
4. Santarsiero, B.D., D.T. Yegian, C.C. Lee, G. Spraggon, J. Gu, D. Scheibe, D.C. Uber, E.W. Cornell, R.A. Nordmeyer, W.F. Kolbe, J. Jin, A.L. Jones, J.M. Jaklevic, P.G. Schultz, and R.C. Stevens, *An approach to rapid protein crystallization using nanodroplets.* J Appl Cryst, 2002. **35**(2): p. 278-281.
5. Gilliland, G.L., M. Tung, and J.E. Ladner, *The Biological Macromolecule Crystallization Database: crystallization procedures and strategies.* Acta Crystallogr D, 2002. **58**: p. 916-920.
6. Rupp, B., *Maximum-likelihood crystallization.* Journal of Structural Biology, 2003. **142**(1): p. 162-169.
7. Kantardjieff, K.A. and B. Rupp, *Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals.* Prot Sci, 2003. **12**(9).
8. Bjellqvist, B., J.C. Sanchez, C. Pasquali, F. Ravier, N. Paquet, S. Frutiger, G.J. Hughes, and D. Hochstrasser, *The focusing positions of polypeptides in immobilized gradients can be predicted from their amino acid sequences.* Electrophoresis, 1993. **14**: p. 1023-1031.
9. Baisnee, P.-F., P. Baldi, S. Brunak, and A.G. Pedersen, *Flexibility of the genetic code with respect to DNA structure.* Bioinformatics, 2001. **17**: p. 237-248.
10. Urquhart, B.L., S.J. Cordwell, and I. Humphrey-Smith, *Comparison of predicted and observed properties of proteins encoded by the genome of Mycobacterium tuberculosis H37Rv.* Biochem Biophys Red Comm, 1998. **253**: p. 70-79.

11. South East Collaboratory for Structural Genomics, *Calculated pI values in C. elegans genome*, http://sgce.cbse.uab.edu/sgce/Pdf/histo_pi.pdf.
12. Peek, M.E. and L.D. Williams, *X-ray crystallography of DNA-drug complexes.* Methods Enzymol, 2001. **340**: p. 282-290.
13. Segelke, B. and B. Rupp, *Beyond the Sparse Matrix Screen: A Web Service for Randomly Generating Crystallization Experiments.* ACA Meeting Series, 1998. **25**: p. 78.
14. Segelke, B.W., *Efficiency analysis of sampling protocols used in protein crystallization screening.* J Crystal Growth, 2001. **232**: p. 553-562.
15. McPherson, A., *Preparation and analysis of protein crystals*. 1982: Wiley, New York.
16. Zimm, B.H., *Applications of the methods of molecular distribution to solutions of large molecules.* J Phys Chem, 1946. **14**: p. 164-179.
17. Farnum, M. and C. Zukoski, *Effect of glycerol on the interactions and solubility og bovine pancreatic trypsin inhibitor.* Biophys J, 1999. **76**: p. 2716-2726.
18. Fine, B.M., A. Lornakin, O.O. Ogun, and G.B. Benedek, *The concentration dependence of the diffusion coefficient for bovine pancreatic trypsin inhibitor: a dynamic light scattering study of a small protein.* Biopolymers, 1996. **28**: p. 2001-2024.
19. Rosenbaum, D.F., P.C. Zamora, and C.F. Zukoski, *Phase behavior of small attractive colloidal particles.* Phys Rev Lett, 1996. **76**: p. 150-153.
20. Rosenbaum, D.F. and C.F. Zukoski, *Protein interactions and crystallization.* J Cryst Growth, 1996. **169**: p. 752-758.
21. George, A. and W.W. Wilson, *Predicting protein crystallization from a dilute solution property.* Acra Crystallogr D, 1994. **50**: p. 361-365.
22. George, A., Y. Chiang, B. Guo, A. Arabshahi, Z. Cai, and W.E. Wilson, *Second virial coefficient as predictor in protein growth.* Methods Enzymol, 1997. **276**: p. 100-110.
23. Bonnete, F., S. Finet, and A. Tardieu, *Second Virial coefficient: variations with lysozyme crystallization conditions.* J Cryst Growth, 1999. **196**: p. 403-414.
24. Beretta, S., G. Chirico, and G. Baldini, *Short-tange interactions of globular proteins at high ionic strengths.* Macromol, 2000. **33**: p. 8663-8670.
25. Bonnete, F., D. Vivares, C. Robert, and N. Colloc'h, *Interactions in solution and crystallization of Asperigillus flavus urate oxidase.* J Cryst Growth, 2001. **232**: p. 330-339.
26. Tardieu, A., S. Finet, and F. Bonnete, *Structure of the macromolecular solutions that generate crystals.* J Cryst Growth, 2001. **232**: p. 1-9.
27. Haas, C., J. Drenth, and W.W. Wilson, *Relation between the solubility of proteins in aqueous solutions and the second virial coefficient of the solution.* J Phys Chem B, 1999. **103**: p. 2808-2811.
28. Haas, C. and J. Drenth, *Understanding protein crystallization on the basis of the phase diagram.* J Cryst Growth, 1999. **196**: p. 388-394.
29. Ruppert, S., S.I. Sandler, and A.M. Lenhoff, *Correlation between the osmotic second virial coefficient and the solubility of proteins.* Biotechnol Prog, 2001. **17**: p. 182-187.
30. Sear, R.P., *Distribution of the second virial coefficients of globular proteins.* Europhys Lett, 2002. **60**: p. 938-944.
31. Haas, C. and J. Drenth, *The interface between a protein crystal and an aqueous solution and its effects on nucleation and crystal growth.* J Phys Chem B, 2000. **104**: p. 368-377.
32. Neal, B.L., D. Asthagiri, and A.M. Lenhoff, *Molecular origins of osmotic second virial coefficients of proteins.* Biophys J, 1998. **75**: p. 2469-2477.

33. Juarez-Martinez, G., C. Garza, R. Castillo, and A. Moreno, *A dynamic light scattering investigation of the nucleation and growth of thaumatin crystals.* J Cryst Growth, 2001. **232**: p. 119-131.
34. Harris, L.J. and A. McPherson, *Crystallization of intact monoclonal antibodies.* Proteins: Struct Funct Genetics, 1995. **23**: p. 285-289.